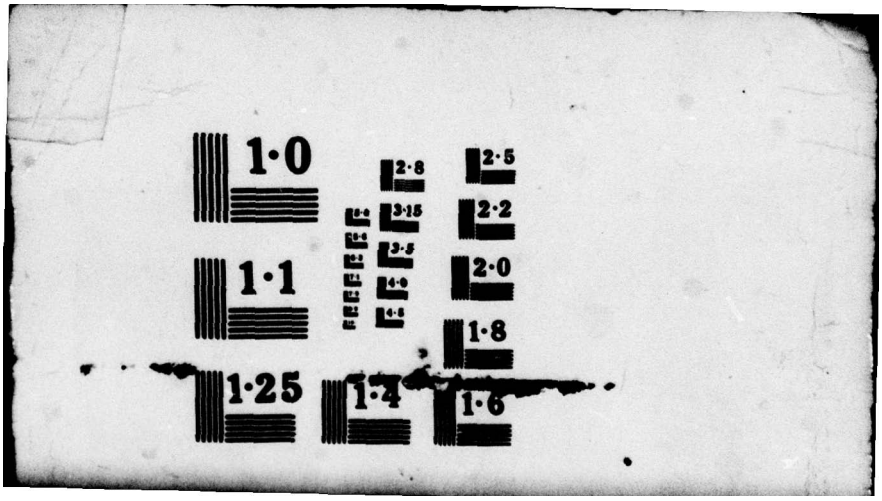
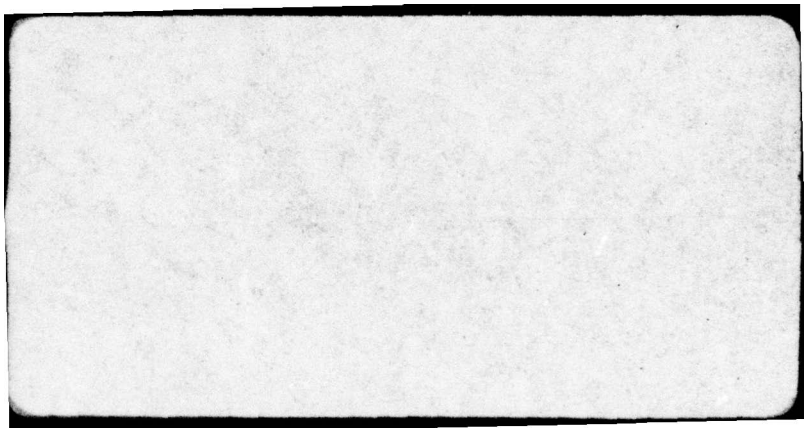
AD-A072 934 SOUTHERN METHODIST UNIV DALLAS TEX DEPT OF STATISTICS F/G 12/1
ON THE RELATIONSHIP BETWEEN THE R AND S ARRAYS AND THE BOX-JENK--ETC(U)
JUL 79 W A WOODWARD, H L GRAY N00014-75-C-0439

UNCLASSIFIED TR-134 NL

END
DATE FILMED
9 79
DDC


1·0
1·1
1·25
1·4
1·6
1·8
2·0
2·2
2·5
2·8
3·15
3·5
4·0
4·5



(12)

DDC
RECEIVED
AUG 21 1979
C

(9) Technical rept.

(6)

# ON THE RELATIONSHIP BETWEEN THE R AND S ARRAYS AND THE BOX-JENKINS METHOD OF ARMA MODEL IDENTIFICATION.

(10)

by

Wayne A. Woodward and H. L. Gray
Southern Methodist University

(14) TR-

Technical Report No. 134
Department of Statistics ONR Contract
(11) 25 July 1979

(12) 39 p.

(15)

Research Sponsored by the Office of Naval Research
Contract N00014-75-C-0439
Project NR 042-280



DIVISION OF MATHEMATICAL SCIENCES
Department of Statistics
Southern Methodist University
Dallas, Texas 75275





404 198

# On the Relationship Between the R and S Arrays and the Box-Jenkins Method of ARMA Model Identification

Wayne A. Woodward and H. L. Gray
Southern Methodist University

## ABSTRACT

In this paper an extension of the partial autocorrelation function which we call the generalized partial autocorrelation function is investigated. This generalized partial autocorrelation function is useful in examining the relationship between the R- and S-array method of Gray, Kelley, and McIntire and the Box-Jenkins approach to ARMA model identification. Also the generalized partial autocorrelation is shown to be a useful model identification tool to be used along with the R- and S-arrays. Also discussed is a reformatting of the S-array into the *Shifted S-array* which the authors believe to be easier to use in practice than the S-array. The methods of this paper are illustrated by means of examples including an analysis of the Makridakis (1978) metals series data.

Key words: autoregressive-moving average processes; generalized partial autocorrelation function; partial autocorrelation function; R- and S-arrays; time series

This research was partially supported under ONR Contract N00014-75-C-0439

# On the Relationship Between the R and S Arrays and the Box-Jenkins Method of ARMA Model Identification

Wayne A. Woodward and H. L. Gray
Southern Methodist University

## 1. INTRODUCTION

Gray, Kelley, and McIntire (1978) have proposed an alternative to the Box-Jenkins approach of ARMA (p,q) model identification based on R- and S-arrays. Their method was shown to perform well in practice, and it uniquely determines p and q when the true autocorrelation function is known. Using the Box-Jenkins approach, however, when the autocorrelation function is known, unique determination of p and q is only assured when either $p = 0$ or $q = 0$.

In this paper the concept of a generalized partial autocorrelation function, a natural extension of the partial autocorrelation function, is discussed. It will be shown that more information concerning the order of an ARMA (p,q) process is available in the generalized partial autocorrelation function than in the partial autocorrelation function, even when $q = 0$. It will also be shown that the generalized partial autocorrelation function can be obtained as a ratio of elements in the S-array, and that some significant information in the S-array can be lost when the ratio is taken.

## 2. METHODS OF MODEL IDENTIFICATION

Consider the ARMA (p,q) process given by

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \qquad (2.1)$$

where $\phi_k$, $k = 1, \ldots, p$ and $\theta_j$, $j = 1, \ldots, q$ are real valued constants with $\phi_p \neq 0$ and $\theta_q \neq 0$ and where $a_t$ is white noise. Employing the backward shift operator B, defined by $BX_t = X_{t-1}$, equation (2.1) is often written in the form $\phi(B)X_t = \theta(B)a_t$ where

$$\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q.$$

It will be assumed that the two equations

$$\phi(r) = 1 - \phi_1 r - \ldots - \phi_p r^p = 0$$

and

$$\theta(r) = 1 - \theta_1 r - \ldots - \theta_q r^q = 0$$

have no common roots.

Letting $\rho_k$ denote the autocorrelation function at lag k, it can be shown that if $X_t$ is a stationary ARMA(p,q) process, then

$$\rho_k - \phi_1 \rho_{k-1} - \ldots - \phi_p \rho_{k-p} = 0, \; k \geq q + 1. \qquad (2.2)$$

In particular, if $q = 0$, then

$$\rho_k - \phi_1 \rho_{k-1} - \ldots - \phi_p \rho_{k-p} = 0, \; k \geq 1. \qquad (2.3)$$

From (2.3) we obtain the well-known Yule-Walker equations, i.e.,

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2 \rho_1 + \ldots + \phi_p \rho_{p-1} \\
\rho_2 &= \phi_1 \rho_1 + \phi_2 + \ldots + \phi_p \rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \ldots + \phi_p \\
\rho_{p+1} &= \phi_1 \rho_p + \phi_2 \rho_{p-1} + \ldots + \phi_p \rho_1 \\
\rho_{p+2} &= \phi_1 \rho_{p+1} + \phi_2 \rho_p + \ldots + \phi_p \rho_2 . \\
&\vdots
\end{aligned} \qquad (2.4)$$

Basic to the Box-Jenkins approach of ARMA model identification is the partial autocorrelation function given by

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & & & & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & & & & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}} \quad . \tag{2.5}$$

That is, if it is assumed that the process actually is ARMA (k,0), regardless what the model actually is, then $\phi_{kk}$ is the Cramer's Rule solution of the first k Yule-Walker equations in (2.4), with k = p, for the last autoregressive coefficient. The Box-Jenkins procedure uses the fact that if X(t) actually is ARMA(p,0) then $\phi_{kk}$ is nonzero for k = p and identically zero for k > p. Also utilized in the Box-Jenkins procedure is the fact that if the process is ARMA (0,q) then $\rho_k = 0$, k > q. That procedure involves the inspection of graphs of the autocorrelation function and partial autocorrelation functions. However, when p and q are both greater than zero, the autocorrelation and partial autocorrelation function do not possess graphs which uniquely determine p and q by simple inspection.

Upon examination of the Yule-Walker Equations in (2.4) it can be seen that if X(t) is ARMA (p,q) then the autocorrelation function does not satisfy the first q Yule-Walker equations but does satisfy equations q + 1 and following. Using this observation we can define the generalized partial autocorrelation (GPAC) function as

$$\phi_{kk}^{(j)} = \frac{\begin{vmatrix} \rho_j & \rho_{j-1} & \cdots & \rho_{j-k+2} & \rho_{j+1} \\ \rho_{j+1} & \rho_j & \cdots & \rho_{j-k+3} & \rho_{j+2} \\ \vdots & & & & \\ \rho_{j+k-1} & \rho_{j+k-2} & \cdots & \rho_{j+1} & \rho_{j+k} \end{vmatrix}}{\begin{vmatrix} \rho_j & \rho_{j-1} & \cdots & \rho_{j-k+2} & \rho_{j-k+1} \\ \rho_{j+1} & \rho_j & \cdots & \rho_{j-k+3} & \rho_{j-k+2} \\ \vdots & & & & \\ \rho_{j+k-1} & \rho_{j+k-2} & \cdots & \rho_{j+1} & \rho_j \end{vmatrix}} . \qquad (2.6)$$

Thus, $\phi_{kk}^{(j)}$ is the Cramer's Rule solution for the last autoregressive coefficient using the $(j + 1)^{st}$ through $(j + k)^{th}$ Yule Walker equations in (2.4) with p = k. That is, $\phi_{kk}^{(j)}$ is the last autoregressive coefficient if it is assumed that the process is ARMA (k,j). It is easily seen that if $X_t$ is actually ARMA (p,q) then $\phi_{pp}^{(q)} = \phi_p$. Also, if $X_t$ is ARMA (p,q) then $\phi_{kk}^{(q)}$ is nonzero for k = p and identically zero for k > p. In addition if the process is ARMA (p,q) one would normally use equations q + 1 through q + p to solve for the parameters. However, one could utilize the p equations q + i through q + i + p - 1 for i = 1, 2, ... and obtain the same solution for the

parameters if the true autocorrelation function is known. Thus, if X(t) is ARMA (p,q) then $\phi_{pp}^{(q+i)} = \phi_p$, i = 0, 1, ... . These properties make it possible to uniquely identify p and q of a mixed process by simple inspection of a table if the true autocorrelation function is known. Granger and Newbold (1977) note that Jenkins has also suggested the use of $\phi_{pp}^{(q)}$ for aiding in model identification.

For each q, q = 0, 1, ... we will consider the GPAC function $\phi_{pp}^{(q)}$, p = 1, 2, ... . Of course $\phi_{pp}^{(0)}$, p = 1, 2, ... is the usual partial autocorrelation function. Two methods of summarizing this collection of generalized partial autocorrelation functions for purposes of identifying p and q will be discussed. A first presentation is the array given in Table 1.

TABLE 1

GPAC Array

| | | Autoregressive Order | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | ... | k | ... |
| Moving Average Order | 0 | $\phi_{11}^{(0)}$ | $\phi_{22}^{(0)}$ | ... | $\phi_{kk}^{(0)}$ | ... |
| | | $\phi_{11}^{(1)}$ | $\phi_{22}^{(1)}$ | ... | $\phi_{kk}^{(1)}$ | ... |
| | | ⋮ | ⋮ | | ⋮ | |
| | ℓ | $\phi_{11}^{(\ell)}$ | $\phi_{22}^{(\ell)}$ | | $\phi_{kk}^{(\ell)}$ | ... |
| | | ⋮ | ⋮ | | ⋮ | |

If $X_t$ is an ARMA (p,q) process, then the associated array will have the pattern shown in Table 2.

TABLE 2

GPAC Array for an ARMA (p,q) Process

| Moving Average Order | Autoregressive Order: 1 | ... | p-1 | p | p+1 | p+2 | ... |
|---|---|---|---|---|---|---|---|
| 0 | $\phi_{11}^{(0)}$ | ... | $\phi_{p-1,p-1}^{(0)}$ | $\phi_{pp}^{(0)}$ | $\phi_{p+1,p+1}^{(0)}$ | $\phi_{p+2,p+2}^{(0)}$ | ... |
| . | . | | . | . | . | . | |
| . | . | | . | . | . | . | |
| . | . | | . | . | . | . | |
| q-1 | $\phi_{11}^{(q-1)}$ | ... | $\phi_{p-1,p-1}^{(q-1)}$ | $\phi_{pp}^{(q-1)}$ | $\phi_{p+1,p+1}^{(q-1)}$ | $\phi_{p+2,p+2}^{(q-1)}$ | ... |
| q | $\phi_{11}^{(q)}$ | ... | $\phi_{p-1,p-1}^{(q)}$ | $\phi_{p}$ | 0 | 0 | ... |
| q+1 | $\phi_{11}^{(q+1)}$ | ... | $\phi_{p-1,p-1}^{(q+1)}$ | $\phi_{p}$ | u* | u | ... |
| . | . | | . | . | . | . | |
| . | . | | . | . | . | . | |
| . | . | | . | . | . | . | |

*u = undefined

Thus the procedure is to search for a column p in which constant behavior occurs and a row q in which the elements are zero for columns k, k > p. An alternative to this approach is a graphical procedure in which each row of Table 2 is graphed on the same set of axes and the patterns of Table 2 are viewed graphically. This approach is similar to the Box-Jenkins graphical approach.

Gray, Kelley, and McIntire (1978) approach the problem of identifying p and q by defining R and S array elements as the following ratios:

$$R_n(f_m) = H_n(f_m)/H_n(1;f_n)$$
$$S_n(f_m) = H_{n+1}(1;f_m)/H_n(f_m) \qquad (2.7)$$

where

$$H_n(f_m) = \begin{vmatrix} f_m & f_{m+1} & \cdots & f_{m+n-1} \\ f_{m+1} & f_{m+2} & \cdots & f_{m+n} \\ \vdots & & & \\ f_{m+n-1} & f_{m+n} & \cdots & f_{m+2n-2} \end{vmatrix}$$

and

$$H_{n+1}(1;f_m) = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ f_m & f_{m+1} & & f_{m+n} \\ f_{m+1} & f_{m+2} & & f_{m+n+1} \\ \vdots & & & \\ f_{m+n-1} & f_{m+n} & \cdots & f_{m+2n-1} \end{vmatrix}.$$

In their work, $f_m = \rho_m$ or $f_m = (-1)^m \rho_m$ where $\rho_m$ is the autocorrelation function of the ARMA (p,q) process at lag m. A simple iterative method of calculating the R and S array elements is also available. (See [4]).

The properties of the S-array upon which the Gray, Kelley, and McIntire procedure depend are summarized in Theorem 1.

Theorem 1. Let $X_t$ be a stationary ARMA(p,q) process (p > 0) with autocorrelation $\rho_m$. Suppose $S_n(\rho_m)$ is defined and nonzero for all m.

(a) For some integer $m_0$ and some constant $C_1 \neq 0$,

$S_n(\rho_m) = C_1$, $m \geq m_0$

$S_n(\rho_{m_0-1}) \neq C_1$

iff $n = p$ and $m_0 = q - p + 1$.

Moreover $C_1 = (-1)^p \left[1 - \sum_{k=1}^{p} \phi_k\right]$.

(b) Suppose $S_n(\rho_m)$ is defined and nonzero for all m.

For some integer $m_1$ and constant $C_2 \neq 0$,

$S_n(\rho_m) = C_2$, $m \leq m_1$

$S_n(\rho_{m_1+1}) \neq C_2$

iff $n = p$ and $m_1 = -q - p$.

Moreover $C_2 = -C_1/\phi_p$.

(c) For $k > n$, $S_k(\rho_{-k-m}) = \pm \infty$

and $S_k(\rho_{-k+m+1}) = (-1)^{k-n} S_n(\rho_{-n+m+1})$

iff $n = p$ and $m = q$.

Proof: Proofs for (a) and (b) are given by Gray, Kelley, and McIntire (1978) and in more detail by McIntire (1977). The proof of (c) follows easily from (2.7).

The S values are placed in an S-array as in Table 3 where we have employed the shortened notation $S_n(m) = S_n(f_m)$.

TABLE 3

| m \ n | 1 | ... | k |
|---|---|---|---|
| $-\ell$ | $S_1(-\ell)$ | ... | $S_k(-\ell)$ |
| $-\ell+1$ | $S_1(-\ell+1)$ | ... | $S_k(-\ell+1)$ |
| ⋮ | ⋮ | | |
| -1 | $S_1(-1)$ | ... | $S_k(-1)$ |
| 0 | $S_1(0)$ | ... | $S_k(0)$ |
| 1 | $S_1(1)$ | ... | $S_k(1)$ |
| 2 | | | |
| ⋮ | ⋮ | | |
| j | $S_1(j)$ | ... | $S_k(j)$ |

In Table 4 we present the behavior of the S-array when $X_t$ is ARMA(p,q) and $f_m = \rho_m$.

TABLE 4

$(S(\rho_m))$

| m \ n | 1 | ... | p† | p+1 | p+2 ... |
|---|---|---|---|---|---|
| $-\ell$ | $S_1(-\ell)$ | ... | $C_2$ | u | u |
| $-\ell+1$ | $S_1(-\ell+1)$ | | $C_2$ | u | u |
| ⋮ | ⋮ | | ⋮ | ⋮ | ⋮ |
| $-q-p-2$ | $S_1(-q-p-2)$ | | $C_2$ | u | u<br>$\pm\infty$ |
| $-q-p-1$ | $S_1(-q-p-1)$ | ... | $C_2$ | $\mp\infty$ | 2q non-constants |
| $-q-p$ | $S_1(-q-p)$ | ... | $C_2$ | | |
| ⋮ | ⋮ | | 2q non-constants | 2q non-constants | $(-1)^2 C_1$ |
| $+q-p$ | $S_1(q-p)$ | ... | | $-C_1$ | u |
| $q-p+1$ | $S_1(q-p+1)$ | ... | $C_1$ | u | u |
| ⋮ | ⋮ | | ⋮ | ⋮ | ⋮ |
| j | $S_1(j)$ | ... | $C_1$ | u | u |

†Note constant stretches

The following theorem establishes the relationship between the S-arrays and the generalized partial autocorrelation function.

Theorem 2. Let $\rho_k$ be the autocorrelation function of a stationary time series. Then

$$\phi_{kk}^{(j)} = - S_k(\rho_{-k+j+1}) / S_k(\rho_{-k-j}) .$$

Proof: We will let $|A|$ and $|B|$ denote the numerator and denominator determinants respectively in the definition of $\phi_{kk}^{(j)}$ in (2.6). By definition,

$$\frac{S_k(\rho_{-k+j+1})}{S_k(\rho_{-k-j})} = \frac{H_{k+1}(1;\rho_{-k+j+1})H_k(\rho_{-k-j})}{H_k(\rho_{-k+j+1})H_{k+1}(1;\rho_{-k-j})}$$

$$= (-1)^k H_{k+1}(1;\rho_{-k-j})$$

Also, it can be shown that

$$H_{k+1}(1;\rho_{-k+j+1}) = (-1)^k H_{k+1}(1;\rho_{-k-j})$$

In addition

$$H_k(\rho_{-k+j+1}) = (-1)^{[k/2]}|B| \text{ and } H_k(\rho_{-k-j}) = (-1)^{[\frac{k-1}{2}]}|A|$$

where [ ] denotes the greatest integer function. Thus

$$\frac{-S_k(\rho_{-k+j+1})}{S_k(\rho_{-k-j})} = \frac{(-1)(-1)^k(-1)^{[\frac{k-1}{2}]}|A|}{(-1)^{[\frac{k}{2}]}|B|}$$

$$= \frac{|A|}{|B|}$$

$$= \phi_{kk}^{(j)}.$$

As a practical point it should be noted that although the elements of the GPAC array could be calculated via their defining relation (2.6), their calculation is facilitated using Theorem 2 and the iterative procedure for calculating S-array elements.

Before proceeding further it should be pointed out that a result similar to that of Theorem 1 holds when $f_m = (-1)^m \rho_m$ (see [4]). Gray, Kelley, and McIntire recommend the inspection of the S-arrays both with $f_m = \rho_m$ and with $f_m = (-1)^m \rho_m$ since the model identification behavior in the presence of noise is often clearer in one of the arrays than in the other, yet neither is uniformly better than the other. It does appear that for low frequency data it is better to take $\hat{f}_m = (-1)^m \hat{\rho}_m$ in computing Table 3 and for high frequency data S-arrays are usually better at $\hat{f}_m = \hat{\rho}_m$ where $\hat{\rho}_m$ will be defined in Section 3.

Theorem 3 establishes the relationship between the generalized partial autocorrelation function and the S-array with $f_m = (-1)^m \rho_m$.

Theorem 3: Let $\rho_k$ be the autocorrelation function of a stationary time series $X_t$ and let $f_m = (-1)^m \rho_m$. Then,

$$\phi_{kk}^{(j)} = (-1)^{k+1} S_k(f_{-k+j+1}) / S_k(f_{-k-j}) .$$

Proof: It is easily shown, that when $f_m = (-1)^m \rho_m$,

$$\frac{- S_k(f_{-k+j+1})}{S_k(f_{-k-j})} = \frac{|A^*|}{|B^*|}$$

where $|A^*|$ and $|B^*|$ are defined as were $|A|$ and $|B|$ in the proof of Theorem 2 with $\rho_m$ replaced by $(-1)^m \rho_m$. Through row and column operations on $|A^*|$ and $|B^*|$, it can be shown that $|A^*|/|B^*| = (-1)^k |A|/|B|$ and the result follows.

Theorem 3 shows that the GPAC array can be calculated from the S-array at $f_m = \rho_m$ or $f_m = (-1)^m \rho_m$.

## 3. MODEL IDENTIFICATION USING THE GPAC ARRAY- ILLUSTRATIVE EXAMPLES

In this section we will illustrate the use of the GPAC array through three examples. In these examples the autocorrelation function will be estimated based upon a realization of length T by

$$\hat{\rho}_m = \frac{\sum_{t=0}^{T-m} (X_t-\bar{X})(X_{t+m}-\bar{X})}{\sum_{t=0}^{T} (X_t-\bar{X})^2} .$$

It will be seen that the patterns in the GPAC array, although somewhat disturbed, are still very useful in identifying the order of the model. Example 3 indicates a caution which must be exercised in the use of the GPAC array for determining the moving average order.

Finally, before proceeding to the examples it should be mentioned that if p = 0, the Box-Jenkins procedure, the Gray, Kelley, McIntire procedure, and the GPAC procedure all must utilize the autocorrelation function with its property that $\rho_k = 0$, $k \geq q + 1$.

Example 1. Consider the process

$$X_t - 1.34\, X_{t-1} + .65\, X_{t-2} = a_t. \qquad (3.1)$$

In Table 5 the GPAC array using the true autocorrelation function $\rho_m$ is presented. The identification as an ARMA (2,0) process is clear due to the fact that $\phi^{(0)}_{2+i,2+i} = 0$, i = 1, 2, ... and that $\phi^{(i)}_{22} = -.65$, i = 0, 1, 2, ... . In Table 6 the sample generalized partial

autocorrelation function array is given based on a realization from (3.1) of length 100. The patterns which occurred exactly in Table 5 are still visible in Table 6, i.e., $|\hat{\phi}^{(0)}_{2+1,2+1}| \leq .127$, $i = 1, \ldots 8$ and $\hat{\phi}^{(i)}_{22} \approx -.65$, $i = 0, \ldots 5$. Of course the process is strictly autoregressive and it is clear that one could have identified the process as ARMA (2,0) strictly on the basis of the first row of Table 6 which is the partial autocorrelation function. However, as will usually be the case, the constant behavior, which occurred in the second column in this example, is the most visible pattern. In any event the combination of both is clearly more informative than either alone. In Example 3 we will deal more completely with the comparison of the model identification capabilities of the constant column behavior and the zero row behavior in the GPAC array.

TABLE 5

True Generalized Partial Autocorrelation Function Array for Series (3.1)

| | | Autoregressive Order | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| | 0 | .812 | -.650 | .000 | .000 | .000 | .000 | .000 | .000 |
| | 1 | .540 | -.650 | u* | u | u | u | u | u |
| Moving | 2 | .135 | -.650 | u | u | u | u | u | u |
| Average | 3 | -3.458 | -.650 | u | u | u | u | u | u |
| Order | 4 | 1.528 | -.650 | u | u | u | u | u | u |
| | 5 | .915 | -.650 | u | u | u | u | u | u |

*u = undefined

TABLE 6

Sample Generalized Partial Autocorrelation Function Array for a Realization of Length 100 from Series (3.1)

| | | Autoregressive Order | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| | 0 | .821 | -.650 | .029 | -.114 | -.075 | -.127 | -.034 | .091 |
| | 1 | .562 | -.628 | -2.487 | -.133 | .116 | -.107 | -.370 | .074 |
| Moving | 2 | .209 | -.702 | .392 | .232 | -.196 | -.170 | -.160 | .012 |
| Average | 3 | -1.894 | -.715 | .873 | 1.143 | -.738 | -.278 | -.167 | -2.244 |
| Order | 4 | 1.901 | -.670 | -.276 | -1.067 | .450 | -.561 | -.343 | .481 |
| | 5 | 1.195 | -.720 | 3.137 | -1.292 | -1.692 | -.281 | .480 | -.052 |

Example 2. Consider the process

$$X_t - 1.5\,X_{t-1} + 1.21\,X_{t-2} - .455\,X_{t-3} = a_t + .2a_{t-1} + .9a_{t-2}. \quad (3.2)$$

In Tables 7 and 8 are given the GPAC arrays based on theoretical autocorrelations and sample autocorrelations from a realization of size 300 respectively. In Table 7 the fact that $\phi^{(2)}_{3+i,3+i} = 0$, $i = 1, 2, \ldots$ and that $\phi^{(2)}_{33} = \phi^{(2+i)}_{33} = .455$, $i = 1, 2, \ldots$ indicates that $p = 3$ and $q = 2$. These patterns again are also quite clear in Table 8.

TABLE 7

True Generalized Partial Autocorrelation Function Array for Series (3.2)

| | | Autoregressive Order | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| | 0 | .845 | -.706 | .414 | .299 | -.304 | -.145 | .245 | .062 |
| | 1 | .606 | -.458 | .836 | .683 | -.434 | -.646 | .279 | .848 |
| Moving | 2 | .391 | -.070 | .455 | .000 | .000 | .000 | .000 | .000 |
| Average | 3 | .328 | 2.073 | .455 | u* | u | u | u | u |
| Order | 4 | 1.356 | -.119 | .455 | u | u | u | u | u |
| | 5 | 1.632 | 5.367 | .455 | u | u | u | u | u |

*u = undefined

TABLE 8

Sample Generalized Partial Autocorrelation Function Array
for a Realization of Length 300 from Series (3.2)

| | | Autoregressive Order | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| | 0 | .819 | -.725 | .344 | .315 | -.164 | -.177 | .145 | .105 |
| | 1 | .528 | -.540 | .930 | .477 | -.494 | -.307 | .270 | .271 |
| Moving | 2 | .165 | -.309 | .458 | -.017 | .101 | .014 | .004 | .037 |
| Average | 3 | -1.124 | -.204 | .452 | 2.850 | .104 | -.017 | -.123 | .045 |
| Order | 4 | .296 | -1.048 | .468 | .024 | .092 | -1.702 | -.316 | .072 |
| | 5 | -4.174 | -1.596 | .467 | -1.771 | .145 | .005 | .294 | .393 |

Example 3. In Table 9 is presented the GPAC array using the true autocorrelation function for the process

$$X_t - .19\, X_{t-1} + .75\, X_{t-2} = a_t - .5\, a_{t-1}. \qquad (3.3)$$

In that array it should be noted that

$$\phi_{22}^{(j)} = -.75,\ j \geq 0 \quad \text{and} \quad \phi_{kk}^{(1)} = 0,\ k \geq 2\,.$$

Since the process is ARMA (2,1) we would have expected $\phi_{22}^{(j)} = -.75$, $j \geq 1$ and $\phi_{22}^{(0)} \neq -.75$. This example, however, points out that the constant behavior in the $p^{th}$ column of the GPAC array of a stationary ARMA (p,q) process may begin prior to row q. In other words if the process actually is ARMA (p,q), with $\phi_p$ the $p^{th}$ autoregressive coefficient, and it is treated as an ARMA (p,q-1), for example, the estimate of the $p^{th}$ autoregressive coefficient in this case may also be $\phi_p$. Thus, the user must exercise caution when

using the GPAC array to determine q. The constant behavior in a column of the GPAC array should be accompanied by the appropriate zero behavior. It is clear that for a stationary ARMA process, $\phi_{kk}^{(q)} = 0$ for $k \geq p + 1$ and $\phi_{pp}^{(q)} \neq 0$ if and only if the process is ARMA (p,q). Thus, in the array in Table 9, although the constant behavior is misleading, the zero behavior correctly identifies p and q. We will formally state the results of the above discussion in the following theorem.

Theorem 4. Let $X_t$ be a stationary ARMA(p,q) process (p > 0).

(a) $\phi_{pp}^{(j)} = \phi_p$, $j \geq q$.

(b) Suppose that the nth order Yule-Walker equations are non-singular. For some constant m, $\phi_{kk}^{(m)} = 0$, $k \geq n+1$ and $\phi_{nn}^{(m)} \neq 0$ iff $n = p$ and $m = q$.

Referring back to Theorem 1 and the corresponding results for $f_k = (-1)^k \rho_k$ (see [4]) it is seen that the constant behavior in the S-array is necessary and sufficient. In Table 10 is presented the 2nd column of the S-array with $f_k = \rho_k$ from which the GPAC array in Table 9 was calculated. We see that, as would be expected from Theorem 1, the ambiguity seen in Table 9 is not present in Table 10. The additional constant term in Table 9 was due to the fact that 1.749/2.333 = 1.560/2.080 to 3 decimal places. It is obvious that sequences of constants in the GPAC array due to constant ratios with numerators and denominators which vary, can in fact occur in any column of the GPAC array. Thus, the S-array should always be checked when considering constant behavior in the GPAC array. In Section 5 we will demonstrate that the zero behavior of GPAC can also be misleading. This problem is also alleviated via the S-array as we will see.

TABLE 9

Theoretical GPAC Array for Series (3.3)

| | | Autoregressive Order | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| | 0 | .000 | -.750 | -.325 | -.157 | -.078 | -.039 | -.019 | -.010 |
| | 1 | -2139.989 | -.750 | -.000 | .000 | .000 | .000 | .000 | .000 |
| Moving | 2 | .190 | -.750 | u* | u | u | u | u | u |
| Average | 3 | -3.750 | -.750 | u | u | u | u | u | u |
| Order | 4 | .390 | -.750 | u | u | u | u | u | u |
| | 5 | -1.733 | -.750 | u | u | u | u | u | u |

*u = undefined

TABLE 10

Column 2 from Theoretical S-Array with

$f_k = \rho_k$ for Series (3.3)

| Lag | Column 2 |
|---|---|
| -7 | 2.587 |
| -6 | 2.587 |
| -5 | 2.587 |
| -4 | 2.587 |
| -3 | 2.587 |
| -2 | 2.334* |
| -1 | 1.751* |
| 0 | 1.940 |
| 1 | 1.940 |
| 2 | 1.940 |
| 3 | 1.940 |
| 4 | 1.940 |

## 4. THE SHIFTED S-ARRAY

The present investigation of the GPAC array has led to a reexamination of the S-array. For a stationary ARMA (p,q) process, the constant behavior in the $p^{th}$ column of the GPAC array relates closely to the constants $C_1$ and $C_2$ in the $p^{th}$ column of the S-array. The "zero" behavior in the GPAC array in turn corresponds to Theorem 1(c). That is if $X_t$ is ARMA (p,q) then for i = 1, 2, ... we have $S_{p+i}(\rho_{-p-i-q}) = \pm \infty$ and $S_{p+i}(\rho_{-p-i+q+1}) = (-1)^i C_1$ while $\phi^{(q)}_{p+i,p+i} = - S_{p+i}(\rho_{-p-i+q+1})/S_{p+i}(\rho_{-p-i-q}) = 0$. Gray, Kelley, and McIntire (1978) recommend utilizing the information in column p + 1 in the identification of an ARMA (p,q) process, but they do not suggest utilizing the behavior described in Theorem 1(c) for i > 1. In fact the D-statistic (see [4]) utilizes columns p and p + 1 only (in addition to column p + 1 of the R-array). Thus the S-array approach uses the constant behavior and the first zero in the row behavior of the GPAC. It seems that one is in fact ignoring useful model identification information by not considering the behavior in the columns past the $p + 1^{st}$ columns of the S-array. The authors are currently considering a modification of the D-statistic with these comments in mind. This point will be commented on further in Section 5.

The S-array, however, is not presently designed to easily facilitate such an examination of the Theorem 1(c) behavior as one must proceed diagonally through the array to examine this behavior. To alleviate this problem, a modification of the S-array, to be called the Shifted S-array will be introduced. This Shifted S-array

was first proposed by G. Kelley in 1977. Letting $S_k^*(f_j)$ denote the element in the $k^{th}$ column at lag j of the Shifted S-array, we have

$$S_k^*(f_j) = S_k(f_{j-k+1}).$$

Thus, Theorem 1 can be restated as follows.

Theorem 1': Let $X_t$ be a stationary ARMA(p,q) process (p > 0) with autocorrelation $\rho_m$. Suppose that $S_n^*(\rho_m)$ is defined and nonzero for all m.

(a) For some integer $m_0$ and some constant $C_1 \neq 0$,

$S_n^*(\rho_m) = C_1$, $m \geq m_0$

$S_n^*(\rho_{m_0-1}) \neq C_1$

iff $n = p$ and $m_0 = q$.

Moreover $C_1 = (-1)^p[1 - \sum_{k=1}^{p} \phi_k]$.

(b) For some integer $m_1$ and constant $C_2 \neq 0$,

$S_n^*(\rho_m) = C_2$, $m \leq m_1$

$S_n^*(\rho_{m_1+1}) \neq C_1$

iff $n = p$ and $m_1 = -q - 1$.

Moreover $C_2 = -C_1/\phi_p$.

(c) For $k > n$, $S_k^*(\rho_{-m-1}) = \pm \infty$ and

$S_k^*(\rho_m) = (-1)^{k-n} S_n^*(\rho_m)$

iff $n = p$ and $m = q$.

Thus, if $X_t$ is a stationary ARMA (p,q) process, then in columns $p + i$, $i = 1, 2, \ldots$ the value $\pm \infty$ will occur at lag $-q-1$ and $(-1)^i C_1$ will occur at lag q. In addition to simplifying the Theorem 1(c) behavior, Theorem 1' and Table 11 indicate that the constant behavior will also be simplified. In the Shifted S-array for a stationary ARMA (p,q) process the 2q non-constant terms will always be the elements from -q through q-1 regardless of p. That is, the nonconstant behavior will always be centered around lags -1 and 0. This will simplify the use of the information in the S-array and will eliminate the need for the "starred quantities" to assist in locating

**TABLE 11**

**Shifted S-Array with $f_m = \rho_m$**

**where $X_t$ is ARMA (p,q)**

| m \ n | 1 | . . . | p | p+1 | ... | p+i ... |
|---|---|---|---|---|---|---|
| . | | | . | . | | . |
| . | | | . | . | | . |
| . | | | . | . | | . |
| -q-2 | | | $C_2$ | u | | u |
| -q-1 | | | $C_2$ | $\pm\infty$ | ... | $\pm\infty$ |
| -q | . | | | | | |
| . | . | | | | | |
| . | . | | | | | |
| -1 | $S_1^*(-1)$ | Non-constants | | Non-constants | | Non-constants |
| 0 | $S_1^*(0)$ | | | | | |
| 1 | $S_1^*(+1)$ | | | | | |
| . | | | | | | |
| . | | | | | | |
| . | | | | | | |
| q-1 | | | | | | |
| q | . | | $C_1$ | $-C_1$ | ... | $(-1)^i C_1$ |
| . | . | | | | | |
| q+1 | . | | $C_1$ | u | | u |
| q+2 | | | $C_1$ | u | | u |
| . | | | . | . | | . |
| . | | | . | . | | . |
| . | | | . | . | | . |

the correct position for inspection of a particular column. In Table 12 is presented the theoretical Shifted S-array for the ARMA (3,2) process of Example 2 while in Table 13 is presented the Shifted S-array for the realization of the length 300 upon which the GPAC array in Table 8 is based. From Table 12, the $\pm \infty$ behavior occurs at lag $-q-1 = -3$ beginning in column 4. In Table 13 this behavior is manifested in the "large" numbers 268.053, -42.435, -319.257. At lag $q = 2$ in columns 4, 5, and 6 the quantities 4.494, -4.300, and 4.454 respectively correspond to the $(-1)^i C_1$ behavior of Theorem 1(c). These observations along with the two sets of near constants in column 3 of Table 13 identify the process as an ARMA (3,2). Of course the nonconstant behavior is centered at the line drawn between lags 0 and -1. It is the opinion of the authors that the Shifted S-array presents the information in the S-array in a format which is easier to use in practice.

## 5. A COMPREHENSIVE EXAMPLE

In this section we employ the model identification techniques mentioned in this paper, namely the GPAC array and the Shifted S-array to model the Makridakis (1978) metal series data. It is hoped that this example will demonstrate the use of these techniques in modeling, rather easily, a series which has been difficult to handle with previous techniques.

The metal series data consists of 144 monthly values of carbon steel monthly shipments from 1961-1972. Makridakis has fit the model

$$(1 - B)\, Y_t = (1 - \theta_1 B)(1 - \theta_{12} B^{12})\, a_t, \tag{5.1}$$

where $Y_t = X_t - 3082$ and $X_t$ is the data. He reports that this model

TABLE 12

Shifted S-Array for Model (3.2)

| Lag | $S^*_1$ | $S^*_2$ | $S^*_3$ | $S^*_4$ | $S^*_5$ | $S^*_6$ |
|---|---|---|---|---|---|---|
| -8 | -2.872 | | | | | |
| -7 | -2.037 | 3.552 | | | | |
| -6 | -1.613 | 1.247 | -9.154 | | | |
| -5 | -1.737 | 23.860 | -9.154 | u | | |
| -4 | -4.052 | 2.915 | -9.154 | u | u | |
| -3 | -3.556 | 22.437 | -9.154 | ±∞ | ±∞ | ±∞ |
| -2 | -2.651 | 5.686 | -7.573 | 1.470 | 6.806 | -7.911 |
| -1 | -2.184 | 4.456 | -10.750 | -10.450 | 7.148 | -65.674 |
| 0 | -1.845 | 3.148 | -4.452 | 3.122 | -2.173 | 2.489 |
| 1 | -1.606 | 2.606 | -6.334 | -1.004 | -3.527 | -1.090 |
| 2 | -1.391 | 1.578 | -4.165 | 4.165 | -4.165 | 4.165 |
| 3 | -1.328 | -6.044 | -4.165 | u | u | |
| 4 | -2.356 | 2.838 | -4.165 | u | | |
| 5 | -2.632 | -6.691 | -4.165 | | | |
| 6 | -1.964 | 3.598 | | | | |
| 7 | -1.534 | | | | | |

TABLE 13

Shifted S-Array for Realization of Length 300 from Model (3.2)

| Lag | $S^*_1$ | $S^*_2$ | $S^*_3$ | $S^*_4$ | $S^*_5$ | $S^*_6$ |
|---|---|---|---|---|---|---|
| -8 | -2.312 | | | | | |
| -7 | -1.628 | 3.467 | | | | |
| -6 | -.760 | 2.386 | -9.505 | | | |
| -5 | -4.375 | 1.347 | -9.472 | -177.715 | | |
| -4 | -.110 | 6.459 | -9.579 | 8.022 | -48.460 | |
| -3 | -7.064 | 6.664 | -9.543 | 268.053 | -42.435 | -319.257 |
| -2 | -2.893 | 5.049 | -7.264 | -1.905 | 10.998 | 2.990 |
| -1 | -2.221 | 4.329 | -12.256 | -9.160 | 14.689 | 16.074 |
| 0 | -1.819 | 3.138 | -4.218 | 2.888 | -2.414 | 2.840 |
| 1 | -1.528 | 2.728 | -6.752 | .908 | -5.435 | .919 |
| 2 | -1.165 | 2.059 | -4.372 | 4.494 | -4.300 | 4.454 |
| 3 | .124 | 1.319 | -4.334 | -22.865 | -5.046 | |
| 4 | -1.296 | 1.412 | -4.438 | 4.211 | | |
| 5 | 3.174 | 3.807 | -4.435 | | | |
| 6 | -2.593 | 3.322 | | | | |
| 7 | -1.762 | | | | | |

does not forecast as well as the random walk model, an apparent paradox since he believes this to be the optimum model. Parzen (1979) also analyzes the data and suggests the solution to the paradox is that the model of Makridakas is not satisfactory. Parzen identifies the AR (2) process

$$Y_t - .44\, Y_{t-1} - .35\, Y_{t-2} = a_t \tag{5.2}$$

as the best choice according to Parzen's (1974) CAT (criterion autoregressive transfer) criterion. The same result is obtained using Akaike's (1974) FPE criterion. However, Parzen notices that the sample spectrum does not agree very well with the spectrum of the AR (2) model (5.2) and that the second choice according to CAT is an AR (13) (The second choice by FPE is AR (3).). Moreover the spectrum obtained from the fitted AR (13) is quite compatible with the sample spectrum. For this reason Parzen prefers the AR (13) model even though it is not selected as best by either CAT or FPE.

A plot of the metals data is given in Figure 1 and the sample autocorrelations are shown in Figure 2. Table 14 shows portions of the shifted S-array for the data with $f_m = (-1)^m \hat{\rho}_m$. The array is evaluated with $f_m = (-1)^m \hat{\rho}_m$ because the data is clearly predominantly low frequency (see Gray, Kelley, McIntire (1978)). For that array the D-statistic suggests that the array be examined for the possibility of an ARMA (1,1), ARMA (1,2), ARMA (1,6) or an ARMA (13,1) or AR (13). Examination of the shifted S-array immediately suggests from columns 1 and 13 that the process is ARMA (13,1) with a first order factor of approximately (1 - .9B). Moreover since the two nonconstant values

Figure 1. Makridakis Metals Data

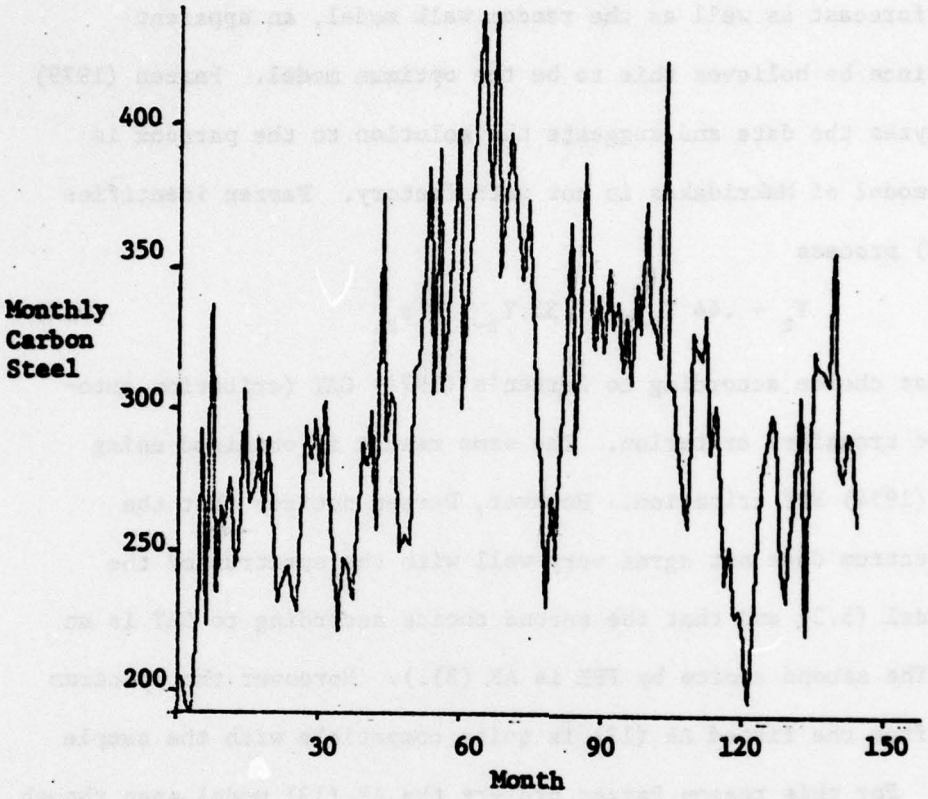
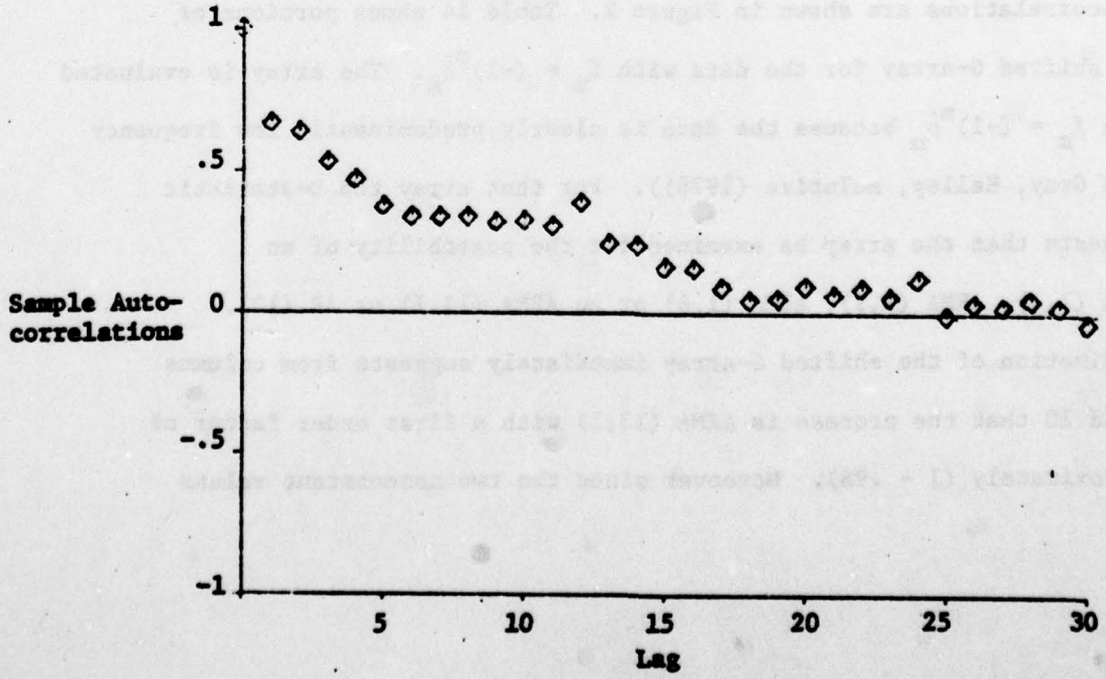



Figure 2. Sample Autocorrelations from the Makridakis Data



in column 13 are close to the "constants" of column 13 it is clear that the moving average term is not large. In keeping with the procedure of Gray, Kelley, and McIntire, the data was transformed by estimating the coefficient of B in $(1 - \phi B)$ by

$$\hat{\phi} = \frac{S_1(1) + S_1(2) + S_1(3)}{S_1(-2) + S_1(-3) + S_1(-4)} \approx .9 .$$

TABLE 14

Portion of Shifted S-Array for
Makridakis Metals Series

$(f_m = (-1)^m \hat{\rho}_m)$

| Lag | $S_1^*$ | $S_2^*$ | $S_3^*$ | ... | $S_{12}^*$ | $S_{13}^*$ | $S_{14}^*$ | $S_{15}^*$ |
|---|---|---|---|---|---|---|---|---|
| -5 | -2.240 | .091 | -.399 | | -3.079 | 2.143 | -5.921 | 4.662 |
| -4 | -2.134 | .406 | -1.313 | | 1.780 | 2.755 | -.400 | 2.006 |
| -3 | -2.206 | -3.458 | 3.731 | | -4.595 | 3.020 | -2.152 | 1.174 |
| -2 | -2.048 | -4.468 | 18.637 | | .969 | 3.263 | -17.066 | 1.676 |
| -1 | -2.493 | -3.168 | -30.088 | ... | -3.951 | 3.008 | 8.151 | 10.565 |
| 0 | -1.670 | 1.093 | -1.135 | | .815 | -.641 | .696 | -.653 |
| 1 | -1.954 | 1.254 | -1.511 | | -.201 | -.953 | 1.156 | -.197 |
| 2 | -1.829 | 1.149 | 5.814 | | 1.236 | -.935 | -1.811 | -.134 |
| 3 | -1.882 | -.424 | .117 | | -.488 | -.731 | -.111 | -1.535 |
| 4 | -1.807 | -.081 | -4.636 | | 1.203 | -.657 | 1.229 | -1.344 |

It is of course more conventional to difference the data in such circumstances as this. However, the data clearly does not indicate a strong tendency toward nonstationarity but does suggest the possibility of a root near $(.9)^{-1}$ in the characteristic equation. Since as has been demonstrated by Gray, Kelley, and McIntire (1978), a factor of the form (1 - .9B) is often close enough to the nonstationary region to produce a nearly singular autocorrelation function, the S-array of $(1 - .9B)X_t$ should be examined as confirmation of the tentative identification of the process as an ARMA (13,1). Although in most problems it makes little difference at this stage of the identification process whether one transforms the data by the difference operator 1 - B or the operator 1 - .9B, these authors have found some cases where this seeming small change in the operators can have a large effect. The shifted S-array of $(1 - .9B)X_t$ is shown in Table 15. The suggested model there is clearly an ARMA (12,1) and the moving average effect is obviously not large. This is consistent with our initial identification of an ARMA (13,1). Thus the identified model is

$$\phi(B)X_t = (1 - \theta_1 B)a_t \tag{5.3}$$

where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_{13} B^{13} .$$

Estimation of these coefficients shows only $\phi_1$, $\phi_{12}$, $\phi_{13}$ and $\theta_1$ to be significantly different from zero and hence suggests the parsimonious model

TABLE 15

Portion of Shifted S-Array for $Z_t = (1-.9B)X_t$

where $X_t$ is the Makridakis Metals Series

$(f_m = (-1)^m \hat{\rho}_m)$

| Lag | $S_1^*$ | $S_2^*$ | $S_3^*$ | ... | $S_{11}^*$ | $S_{12}^*$ | $S_{13}^*$ | $S_{14}^*$ | $S_{15}^*$ |
|---|---|---|---|---|---|---|---|---|---|
| -5 | -.187 | .252 | .657 | | 2.007 | -1.051 | 3.166 | -1.475 | 1.989 |
| -4 | -.264 | 3.037 | 1.534 | | -3.219 | -1.239 | -.300 | -1.746 | -2.014 |
| -3 | 1.298 | -1.789 | .410 | | 4.449 | -1.437 | 1.504 | -1.848 | -6.442 |
| -2 | 2.560 | -7.586 | -.570 | | -1.381 | -1.422 | -43.835 | .258 | 4.163 |
| -1 | 1.188 | 5.875 | 11.809 | | 2.056 | -1.181 | -2.534 | -4.810 | 4.054 |
| 0 | -.543 | .598 | -.569 | | -.397 | .297 | -.337 | .315 | -.292 |
| 1 | -.719 | .853 | .092 | | .220 | .494 | -.480 | -.054 | -.350 |
| 2 | -.565 | -1.311 | -.075 | | -.973 | .496 | -9.085 | .372 | -.387 |
| 3 | .359 | -.252 | -1.514 | | 1.040 | .401 | -.069 | .498 | -.260 |
| 4 | .230 | 4.512 | -.802 | | -1.444 | .356 | -.584 | .485 | -3.975 |

$$(1 - \phi B)(1 - \psi B^{12})Y_t = (1 - \theta B)a_t. \qquad (5.4)$$

Initial estimates for $\phi$, $\psi$ and $\theta$ are

$$\hat{\phi} = .9, \quad \hat{\psi} = .35, \quad \hat{\theta} = .46 .$$

The spectral density of the ARMA (13,1) model fitted from (5.3) is shown in Figure 3. This is essentially the same spectral density as that found by Parzen and as mentioned is quite consistent with the sample spectral density.

Thus far we have not considered the generalized partial autocorrelation in this example. This is to some extent due to the fact that as mentioned before, its column behavior is not sufficient to characterize an ARMA (p,q) process, even though its row behavior does characterize such a process. Nevertheless the generalized partial autocorrelation is quite useful and these authors always consider the GPAC array as well as the S-array. In this way there is no danger in misconstruing the constant column behavior found in the GPAC array. In this example the GPAC presents the model identification quite nicely and at the same time shows why the zero behavior in the rows should not be used alone for identification, even for AR (p) models only. Table 16 shows the GPAC for the metals data and Table 17 for $(1-.9B)X_t$. Row 1 of Table 16 clearly demonstrates why CAT and FPE chose an AR (2) model. That is, since $\phi^{(0)}_{3,3}$ through $\phi^{(0)}_{11,11}$ are all approximately zero it is not surprising that CAT and FPE pick an AR (2). On the other hand note that column 2 in the GPAC is not even approximately constant and hence the long string of zeros in row 1 is only suggestive of a number of zero

Figure 3. Spectral Density for ARMA (13,1)
Model for Makridakis Data

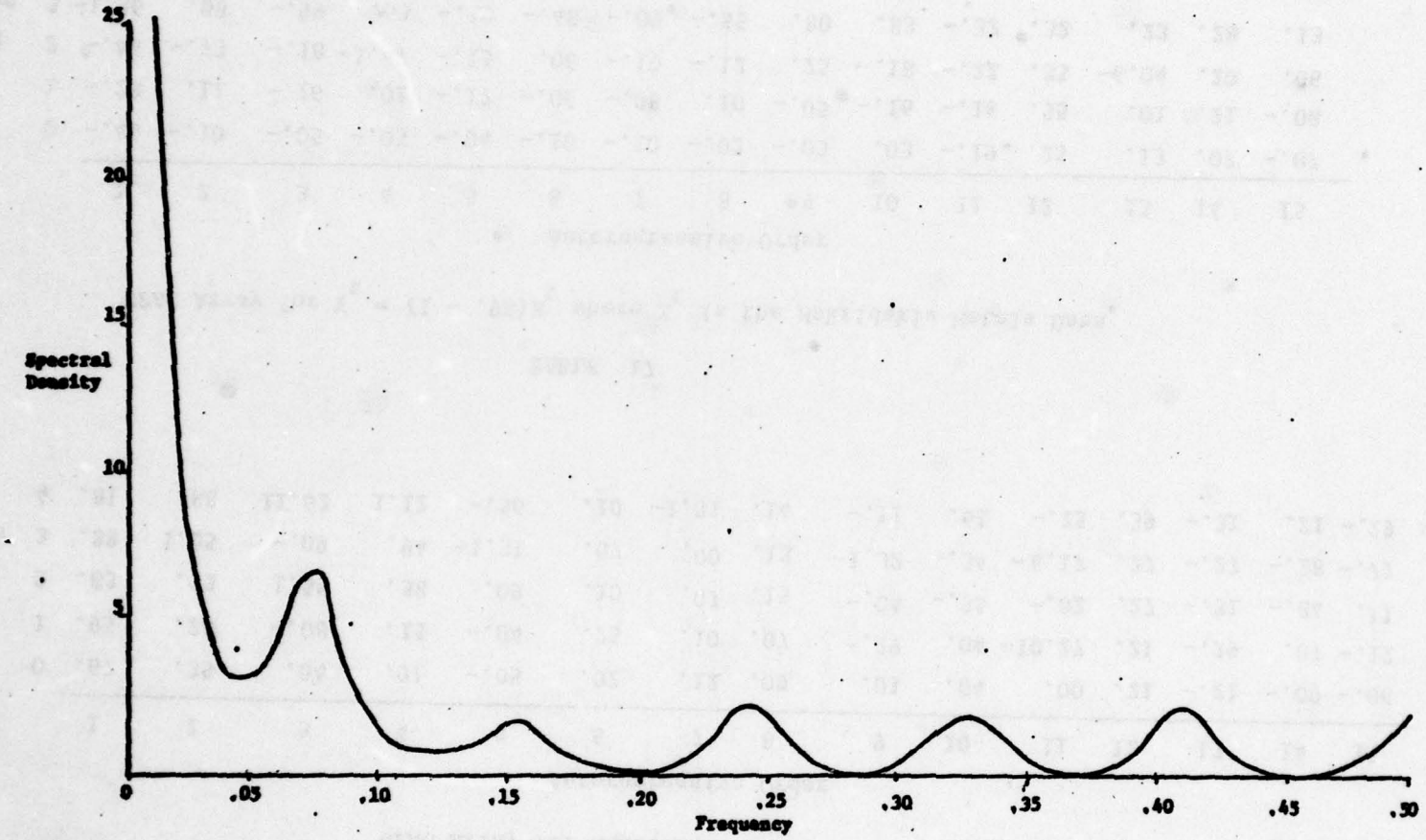


TABLE 16

GPAC Array for Makridakis Data

| | | Autoregressive Order | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| | 0 | .67 | .35 | .04 | .01 | -.05 | .02 | .12 | .09 | .01 | .04 | .00 | .21 | -.21 | -.09 | -.06 |
| | 1 | .95 | .28 | -.08 | .15 | -.04 | .25 | .10 | .07 | -.29 | .04 | -10.27 | .21 | -.29 | .07 | -.12 |
| Moving | 2 | .83 | .33 | 1.56 | .38 | .09 | .10 | .07 | .15 | -.04 | -.35 | -.02 | .27 | -.31 | -.84 | .11 |
| Average | 3 | .88 | 1.05 | -.09 | .94 | -1.31 | .07 | .00 | .13 | -1.32 | -.34 | -6.17 | .27 | -.27 | -.28 | -.77 |
| Order | 4 | .81 | .88 | 11.62 | 1.12 | -.56 | .10 | -1.81 | .14 | -.11 | .61 | -.15 | .39 | -.31 | .21 | -.29 |

TABLE 17

GPAC Array for $Y_t = (1 - .9B)X_t$ where $X_t$ is the Makridakis Metals Data

| | | Autoregressive Order | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| | 0 | -.46 | -.10 | -.05 | -.05 | -.04 | -.10 | -.10 | -.02 | -.03 | .03 | -.19 | .25 | .13 | .07 | -.07 |
| | 1 | -.28 | .11 | -.16 | .01 | -.17 | -.06 | -.08 | .10 | -.05 | -.16 | -.16 | .35 | .01 | .21 | -.08 |
| Moving | 2 | -.44 | -.73 | -.18 | -1.47 | -.15 | .06 | -.10 | -.12 | .25 | .18 | -.22 | .35 | -6.04 | .20 | .06 |
| Average | 3 | -1.36 | .08 | -.99 | .63 | -.19 | -.48 | -.09 | -.55 | .80 | .83 | -.32 | .32 | .23 | .28 | .13 |
| Order | 4 | -1.23 | -17.94 | -1.22 | .20 | .21 | -.15 | .28 | -.13 | -.48 | -.31 | -.72 | .34 | -.19 | .33 | -2.00 |

coefficients prior to $\phi_p$. That is, the row and column behavior together suggest the process is not AR (2) but does have a number of zero coefficients following $\phi_2$. Note this is also clearly seen in the shifted S-array where column 2 shows no constancy behavior. In that array the zero behavior, seen in row 1 of the GPAC array, is obtained from the ratio of the elements $S_k^*(-1)$ and $S_k^*(0)$, i.e. those on each side of the center line. Thus the S-array or shifted S-array and GPAC use both the row and column behavior for identification of a process. The information gained by this procedure is clearly demonstrated in this example where one can easily see that the constant column behavior is influenced by all of the coefficients even if a number of zeros lie between the first and the last autoregressive coefficients. However the zero behavior of rows gives no information as to the values of $\phi_k$ for larger values of k, i.e. those further out than the zeros. The ARMA (13,1) identified by the shifted S-array is also vividly displayed by the GPAC array. Note that the ARMA (13,1) is the only choice which shows both the proper row and column behavior. Nevertheless the large values .954, .826, .882, etc. in row 1 suggests, for reasons already mentioned, that it is a good idea to transform the data to more stationary behavior (as we did) before making the final identification.

In concluding this example we make two final comments. The D-statistic which has been recommended [ 4 ] as an aid in directing the investigator to salient patterns in the S-arrays measures jointly both the constant column behavior of the S-array and the zero row

behavior currently measured by such statistics as CAT and FPE. It is probably overly sensitive to pairwise constant behavior and the very first "zero" in the zero behavior. As mentioned earlier some modifications of this are now underway and it is hoped that it may eventually evolve as more than a guide to inspection of the S-array but as a dependable estimator of p and q for the ARMA (p,q) process. Finally as a word of caution, the GPAC is a very useful measure, however Example 3 which demonstrates that its column behavior is not sufficient to identify q is not pathological and these authors have encountered a number of real data sets where the behavior demonstrated in Example 3 was observed. When these observations are coupled with the present example which demonstrates the zero behavior alone can also be misleading, it is clear that the S-array must be consulted, being the only measure which uniquely characterizes the ARMA (p,q) by both its column and row properties. Actually column and diagonal properties is a better description of this characterization in non-shifted S-arrays. However, believing that shifted S-arrays are easier to "read" for most users, our future reference to S-arrays will mean shifted S-arrays.

## 6. CONCLUDING REMARKS

In this paper we have demonstrated the use of the generalized partial autocorrelation for identifying an ARMA (p,q) model. In the process we have shown that it is the natural extension of the Box-Jenkins method. In addition we have demonstrated the relationship between the S- and R-array approach of Gray, Kelley and McIntire and

the generalized partial autocorrelation. That is, we have shown that the generalized partial autocorrelation, much like the D-statistic, represents a condensation of the information in the S-array. However, it is also shown that the information in the GPAC array can at times be misleading. We thus recommend the use of the GPAC array along with the R- and S-arrays and the D-statistic. Actually we recommend the use of a slight modification of the S-array which we call the shifted S-array because it presents the information in the S-array in a more easily interpretable form.

## REFERENCES


1. Akaike, H. (1970). "Statistical Predictor Identification," Annals of the Institute of Statistical Mathematics 22, 203-217.
2. Box, G.E.P. and Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control. San Francisco: Holden Day, Inc.
3. Granger, C.W.J. and Newbold, P. (1977). Forecasting Economic Time Series. New York: Academic Press.
4. Gray, H.L., Kelley, G.D., and McIntire, D.D. (1978). "A New Approach to ARMA Modeling," Communications in Statistics B7, 1-77.
5. Makridakis, S. (1978). "Time Series Analysis and Forecasting: An Update and Evaluation," International Statistical Review 46, 255-278.
6. McIntire, D.D. (1977). "A New Approach to ARMA Modeling," Ph.D. Dissertation, Southern Methodist University, Dallas, Texas.
7. Parzen, E. (1974). "Some Recent Advances in Time Series Modeling," IEEE Transactions on Automatic Control, AC-19, 723-730.
8. Parzen, E. (1977). "Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes," Multivariate Analysis-IV, edited by P. Krishnaiah, North Holland: Amsterdam, 283-295.
9. Parzen, E. (1979). "Time Series Modeling, Spectral Analysis, and Forecasting," ONR Technical Report No. N-7. Texas A&M Research Foundation, Project No. 3838.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>134 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>ON THE RELATIONSHIP BETWEEN THE R AND S ARRAYS AND THE BOX-JENKINS METHOD OF ARMA MODEL IDENTIFICATION | | 5. TYPE OF REPORT & PERIOD COVERED<br>Technical Report<br>6. PERFORMING ORG. REPORT NUMBER<br>134 |
| 7. AUTHOR(s)<br>Wayne A. Woodward and H. L. Gray | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-75-C-0439 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Southern Methodist University<br>Dallas, Texas 75275 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>NR 042 280 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>OFFICE OF NAVAL RESEARCH<br>ARLINGTON, VA. 22217 | | 12. REPORT DATE<br>July 25, 1979<br>13. NUMBER OF PAGES<br>34 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>UNCLASSIFIED<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

autoregressive-moving average processes; generalized partial autocorrelation function; partial autocorrelation function; R- and S-arrays; time series

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper an extension of the partial autocorrelation function which we call the generalized partial autocorrelation function is investigated. This generalized partial autocorrelation function is useful in examining the relationship between the R- and S-array method of Gray, Kelley, and McIntire and the Box-Jenkins approach to ARMA model identification. Also the generalized partial autocorrelation is shown to be a useful model identification tool to be used along with the R- and S-arrays. Also discussed is a reformatting of the S-array into the Shifted S-array which the authors(see back)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(cont'd)
believe to be easier to use in practice than the S-array. The methods of this paper are illustrated by means of examples including an analysis of the Makridakis (1978) metals series data.